



Facultad de Ciencias Económicas y Empresariales

Sector eléctrico: análisis de la tasa de abandono

Autor: Maria Coronado González

Director: Lourdes Fernández Rodríguez

Resumen

El sector de la energía español es uno de los más complejos. En primer lugar, porque la electricidad no es almacenable y también porque es un SIEG (Servicio de Interés Económico General), lo que implica que es indispensable para el funcionamiento de la sociedad y, por lo tanto, se debe asegurar su disponibilidad.

Desde el inicio el Gobierno interviene en este sector para regularlo y afianzar su funcionamiento. En la actualidad, se encuentra en plena liberalización desde la ley de 1997, con cambios posteriores en la ley de 2013.

Este trabajo de fin de grado analiza la rotación de los clientes que forman parte de la actividad de comercialización de la electricidad. El objetivo principal es el de entender las razones por las que los clientes abandonan la compañía en la que están para irse a otra y que variables determinan la duración de cada cliente. Para ello se utilizan dos modelos de predicción una regresión logística y un *survival analysis*.

Palabras clave: Sector energético, electricidad, tasa de abandono, rotación de clientes, análisis de supervivencia.

Abstract

The Spanish energy sector is one of the most complex in the world. Firstly, because electricity cannot be stored, and secondly because it is an SGEI (Service of General Economic Interest), which means that it is essential for the functioning of society and therefore its availability must be guaranteed.

From the beginning, the government has intervened in this sector to regulate it and strengthen its operation. It is currently under full liberalization since the 1997 law, with subsequent changes in the 2013 law.

This final thesis analyzes the rotation of customers that are part of the electricity commercialization activity. The main objective is to understand the reasons why customers leave the company they are with to go to another one and what affects the duration of each customer. Two predictive models are used for this purpose, a logistic regression, and a survival analysis.

Key words: Energy sector, electricity, churn rate, customer churn, survival analysis.

Tabla de contenido

1. Introducción.....	3
1.1. Estado de la cuestión.....	3
1.2. Motivación.....	5
1.3. Objetivos	5
1.4. Metodología.....	6
1.5. Estructuración del TFG.....	6
2. Marco teórico:	7
2.1. Mercado eléctrico	7
2.2. Evolución de la regulación del mercado eléctrico español	8
2.2.1. Ley 54/1997, de 27 de noviembre, del Sector Eléctrico	9
2.2.2. Ley 24/2013, de 26 de diciembre, del Sector Eléctrico	10
2.3. Estructura y funcionamiento	10
2.3.1. Agentes del mercado eléctrico:	11
2.3.2. Funcionamiento:	13
2.3.3. Tasa de abandono	18
3. Metodología.....	20
3.1. Datos y variables.....	20
3.1.1. Variables originales	21
3.1.2. Variables creadas	23
3.2. Análisis.....	25
3.2.1. Predicción de la tasa de abandono: Regresión	25
3.2.2. Predicción del abandono de los clientes: <i>Survival analysis</i>	30
4. Conclusiones.....	34
5. Bibliografía.....	36
6. Anexos	43

1. Introducción

1.1. Estado de la cuestión

El mercado eléctrico en España ha estado en constante evolución desde la primera vez que se utilizó la electricidad en 1852 para la iluminación de un local en Barcelona. Este hecho fue el comienzo de una serie de cambios y sucesos que dieron lugar a lo que hoy se conoce como la Red Eléctrica Española y el mercado eléctrico no regulado. (Historia de la electricidad en España, 2014).

Pero a lo largo de los últimos años el mercado eléctrico y los precios de este se han visto especialmente afectados por una serie de hechos que han ocurrido en el mundo y que han revolucionado este sector y su funcionamiento. Estos sucesos son la pandemia del Covid-19 y la guerra entre Rusia y Ucrania.

El mercado energético se encontraba en una situación de agitación previa a la pandemia por las discusiones que estaban teniendo lugar entre Rusia y Arabia Saudí por la producción de los combustibles fósiles (Escribano & Lázaro, 2020). Pero todo ello pasó a un segundo plano cuando la pandemia se abrió paso a lo largo de 2020. En el mercado de la energía se preveía un aumento de la demanda en 2020, pero debido a la pandemia lo que se dio fue una reducción de esta a niveles mínimos. Esto sucedió por la paralización global de la actividad económica y de la sociedad en general.

Lo que hace que este mercado sea distinto al resto son las características del producto que comercializa; la energía no se puede almacenar, por lo que se produce para venderla directamente al consumidor.

El mercado eléctrico en España, como se explicará más adelante en este trabajo de fin de grado, funciona según unas curvas de oferta y demanda que para cubrirse utilizan las distintas fuentes de energía disponible. Los precios se ajustan según las curvas del mercado y las predicciones a futuro que se realizan de la demanda y la producción. Como el producto no es almacenable los cambios repentinos en las curvas se reflejan directamente en los precios, alterando el equilibrio del mercado. Esto es lo que sucedió durante la pandemia, se estimó un aumento de la demanda acorde con la tendencia de los meses anteriores a marzo de 2020, pero la realidad es que hubo una reducción y se tuvo que ajustar la curva. En cualquier mercado cuando la demanda disminuye, el precio lo

hace también con el fin de intentar que se venda el producto. En el caso de la electricidad, la reducción de precios fue aún más significativa porque no se puede almacenar y la producción que no se vende se desperdicia (Sánchez, 2022).

El siguiente efecto que tuvo la pandemia fue la recuperación de la actividad económica de forma repentina y disparada. Este hecho implicó un reajuste de la curva de oferta y multiplicó los precios en el mercado eléctrico, lo que se vio afectado de forma directa por el inicio de la guerra entre Rusia y Ucrania en el año 2021 (López, 2022). Este conflicto tiene un efecto a nivel internacional por el alto nivel de exportación de materias primas y otros productos que tienen ambos países. En el caso de Europa el efecto es mayor por la alta dependencia que tiene del petróleo y del gas ruso (Edmon, 2022; Sánchez, 2022). Desde el inicio del conflicto se han impuesto una serie de medidas económicas en los países de la Unión Europea para restringir la importación de productos rusos y de esta forma tratar de limitar la dependencia que tienen de este país. Esto se ha reflejado en un encarecimiento de los precios de la energía y del gas por la consecuente reducción de la oferta del producto y la necesidad de buscar otras fuentes de donde obtenerlo (Sánchez, 2022). Aunque España no importa energía de Rusia, ha sufrido el efecto de la crisis energética que ha provocado la guerra (Azabal, 2022).

En el ámbito nacional, el país se encuentra en plena guerra de comercializadoras y productoras; este tipo de situaciones se han repetido numerosas veces a lo largo de la historia. Se trata de un conflicto en el que cada compañía compite en precios y condiciones para incrementar su cuota de mercado y ganar una ventaja competitiva con respecto al resto. Las guerras de precios son una herramienta que, lejos de fortalecer el mercado, lo destruye (Crampes & Fabra, 2005; Fabra & Toro, 2005). La realidad es que las compañías “se roban” los clientes unas a otras, reduciendo la estabilidad económica del mercado a largo plazo y aumentando la rotación en el sector. Como consecuencia la toma de decisiones y la organización a largo plazo se convierte en una dificultad para las empresas (García, 2022).

Para el correcto funcionamiento del mercado eléctrico es esencial tener la capacidad de realizar predicciones ajustadas a la realidad sobre la demanda y la producción futuras por parte de las empresas generadoras y comercializadoras, porque el producto no es almacenable. Si se produce de más, el exceso se desecha y si se produce de menos,

algunos consumidores no tendrán acceso a la electricidad (Fargas, 2016). El sector energético es complejo e inestable, pero necesita de mucha planificación. Por eso es tan importante que en plena crisis las compañías comercializadoras tengan las herramientas necesarias para realizar predicciones sobre el abandono y la supervivencia de su cartera de clientes, es decir la duración de estos y el momento en que se va cada uno de ellos, para poder hacer predicciones a largo plazo de demanda y ajustar los márgenes a las necesidades futuras.

1.2. Motivación

Los cambios en el mercado energético afectan de forma directa a toda la población, pues como se explicará en el marco teórico (Capítulo 2) se considera un servicio de interés económico general y, por lo tanto, es una necesidad básica para la población. Estos cambios pueden darse por factores coyunturales y temporales, como la pandemia y la guerra entre Ucrania y Rusia, y por causas estructurales e internas, como el desconocimiento por parte de la población del funcionamiento y los componentes del mercado en sí.

Los factores estructurales son los que motivan mi elección de este tema, pues es una de las razones que hay detrás de la guerra de precios entre las comercializadoras y productoras nacionales de energía, como se ha mencionado anteriormente.

Además, este proyecto surge de la realización de unas prácticas en la compañía de la que provienen los datos. En la empresa se observó que el porcentaje de abandono y rotación de los clientes había aumentado significativamente durante los últimos años, lo que inestabilizaba la rentabilidad económica a largo plazo y la predicción a futuro de la demanda. Este hecho creó la necesidad de buscar un modelo con el que poder estimar la tasa de abandono y la duración de la vida de los clientes, y las variables de las que dependen ambos sucesos.

1.3. Objetivos

El principal objetivo de este trabajo es el de buscar una forma de analizar y predecir los clientes que van a abandonar la compañía y cuando lo van a hacer, además de entender la razón por la que este suceso ocurre. Se quiere responder a las siguientes preguntas de investigación: ¿Qué variables afectan más a la decisión de los clientes de dejar la

compañía? ¿Qué variables determinan en mayor medida la duración del contrato de un cliente? Para poder llevar a cabo este análisis es necesario conocer el esquema básico del sector eléctrico y su funcionamiento.

1.4. Metodología

En este trabajo de fin de grado se realiza una revisión bibliográfica para estudiar el sector energético y la información de la que se dispone con respecto al análisis de la tasa de abandono y la duración de los clientes.

Por lo general, el enfoque de la investigación es en su mayoría cuantitativo. Se ha recurrido a técnicas de análisis de datos para determinar el peso de distintas variables sobre la decisión de abandono de los clientes y su tiempo de permanencia en la compañía. Todo ello se lleva a cabo sobre los datos de una de las empresas comercializadoras de la península, por lo que los resultados obtenidos se basan en datos reales, pero por motivos de confidencialidad no se puede citar el nombre de la empresa. Las dos técnicas utilizadas son la regresión logística múltiple y el *survival analysis*. Se han elegido en base a la bibliografía leída y la investigación que existe sobre este y otros sectores en el contexto del abandono de los clientes.

1.5. Estructuración del TFG

La investigación consta de tres partes: la introducción (Capítulo 1) en la que se describe el estado de la cuestión y se pretende poner contexto el resto del estudio, además de estructurarlo y describir la motivación y la metodología del proyecto. El segundo capítulo es el marco teórico, en el que se explica el mercado eléctrico y sus partes; las distintas regulaciones que ha tenido; su funcionamiento y la introducción al abandono de los clientes en el sector. El tercer capítulo estructura el análisis llevado a cabo en tres partes: la explicación de los datos y las variables a estudiar, y los dos métodos utilizados para el análisis de los datos. Ambas técnicas contienen un apartado de explicación del propio modelo y otro de análisis de los resultados. La investigación finaliza con la conclusión obtenida del análisis de los datos y la aplicación de los modelos.

2. Marco teórico:

2.1. Mercado eléctrico

El mercado eléctrico como cualquier otro, proporciona un espacio ordenado y regulado en el que realizar un intercambio de bienes, en este caso electricidad, mediante el cual satisfacer las curvas existentes de oferta y demanda (Rosenbaum, 2000). Este es un mercado internacional e interconectado por la esencialidad del producto que comercializa para la vida actual, pero al mismo tiempo es específico debido a su regulación y al funcionamiento de cada país. A pesar de que el foco de este trabajo está en las empresas comercializadoras de energía y sus clientes, es necesario entender el mercado y su evolución para tener perspectiva a la hora de analizar el abandono y la rotación de los clientes.

En España, la electricidad es un Servicio de Interés Económico General (en adelante SIEG). Este término engloba aquellos servicios cuya disponibilidad se considera indispensable para el funcionamiento de la sociedad, independientemente de si su actividad es rentable para el mercado. Es por esta razón que los sectores que comercializan bienes y servicios considerados un SIEG necesitan la intervención del Estado y regulación que los respalde, para poder garantizar un servicio que es esencial como un derecho para el consumidor (Jimeno & Cebrián, 2015).

El sector eléctrico tiene otras peculiaridades además de funcionar bajo la regularización de los SIEG. Una de las características que más complejidad aporta a este mercado es que el bien que se intercambia en él, la electricidad, no se puede almacenar, al menos no a ese nivel de demanda. Por lo que su funcionamiento tiene que ser perfecto para cubrir la demanda en cada momento y adaptarse a los posibles cambios repentinos que puedan darse en esta. Además existen diferentes formas de producir el bien que se intercambia, pero ninguna es autosuficiente como para cubrir el mercado por completo respetando todas las condiciones de este. Por esta razón la oferta final de electricidad es una combinación de las diversas formas de producción según nivel de contaminación, precio, estabilidad, etc. (Garrués-Irurzun, 2022; Olivares, 2014).

La mayoría de los SIEG tienen altos costes de infraestructura, es decir de inversión inicial, pero también se benefician de economías de escala, lo que define la creación de monopolios naturales y la falta de competencia libre en torno a esos productos y servicios.

En el caso del sector energético, desde 1997 se está llevando a cabo la liberalización del mercado. El principal objetivo de este cambio es conseguir una mayor eficiencia en la determinación de los precios por medio de la competencia entre compañías, porque la regulación de los SIEG no solo busca asegurar la disponibilidad del servicio, sino que también es una medida de protección del consumidor para que este pueda disfrutar de él al menor precio posible, sin sufrir las consecuencias del poder que tienen las compañías líderes en un monopolio para aumentar sus beneficios (Gill de Albornoz & Illueca, 2005; Rodríguez & García, 2017).

2.2. Evolución de la regulación del mercado eléctrico español

A la hora de entender el funcionamiento del mercado es esencial comprender los grandes cambios de regulación que se han dado en este. Como se ha dicho anteriormente el suministro eléctrico es un derecho para el consumidor y, por lo tanto, los organismos encargados de ello tienen la obligación de garantizarlo. Originalmente este propósito se cumplió mediante un control total del sector por parte del Estado, necesidad que surgió de la inestabilidad que presentaba este debido a la guerra civil y las crisis posteriores a las que se enfrentó el país (Velázquez, 2018). Más tarde, se llevó a cabo la liberalización de este, pero manteniendo un mínimo de regulación, pues en ocasiones el mercado no es eficiente por sí solo y se necesita de un creador de mercado para su funcionamiento porque no deja de ser un SIEG (Fargas, 2016).

Antes de la primera ley de 1997 se constituyó el denominado Marco Legal Estable, que era un sistema en el que el Estado intervenía en la determinación de los precios mediante unas cuotas fijas y variables para establecer las tarifas de la electricidad. Después de 1997 las responsabilidades, que hasta ese momento habían sido del Estado, comenzaron a ser de los integrantes del mercado libre y el precio pasó a determinarse a partir de las curvas de oferta y demanda (Llorente, 2016).

2.2.1. Ley 54/1997, de 27 de noviembre, del Sector Eléctrico¹

Las empresas que formaban el mercado antes de la ley de 1997 eran poco numerosas, de gran tamaño y estaban integradas verticalmente, es decir cubrían todos los pasos desde la producción hasta la venta de la electricidad al consumidor final. Todo ello daba lugar a una situación de monopolio natural, lo que dificultaba la liberalización del sector, por lo que uno de los cambios que introdujo esta nueva ley fue la de separar las actividades del sector en cuatro etapas: producción, transporte, distribución y comercialización. De estas cuatro partes, dos de ellas pasaron a funcionar como un mercado de libre competencia (producción y comercialización) y dos siguieron funcionando como monopolio natural con regulación total del Estado (transporte y distribución). Además con la nueva ley se introdujo un quinto agente en la organización del mercado: el operador (Azabal, 2022). Las competencias y funciones de cada uno de los cinco agentes serán desarrolladas más adelante en esta investigación.

La ley no solo cambió la estructuración del mercado, sino que también modificó la composición del esquema de producción de electricidad que se había utilizado hasta ese momento, porque introdujo medidas de apoyo a las plantas de generación de electricidad de origen renovable, cuyo crecimiento en número fue exponencial en los años siguientes (Redondo, 2015).

En los años previos a 1997 hubo diversas guerras de precios, fuente del poder de mercado que tenían las empresas en aquella época. La separación en partes de estas empresas no era suficiente para reducir su poder y su cuota de mercado, por lo que también se crearon los Costes de Transición de Competencia. Su función era cubrir los beneficios que las empresas productoras dejan de ganar por vender su electricidad a un precio competitivo y los costes que los ingresos regulares no soportan (Vives, 2006). Los Costes de Transición de Competencia incluyen dos condiciones: un límite suelo, que define la cantidad máxima que puede obtener una productora, y un límite techo, que determina el precio máximo de venta de la electricidad. De esta forma se buscaba evitar que las

¹ BOE (1997): “Ley 54/1997, de 27 de noviembre, del Sector Eléctrico”. Disponible en: <https://www.boe.es/buscar/doc.php?id=BOE-A-1997-25340> [consulta: 02 de mayo de 2023].

empresas subieran o bajaran los precios según intereses propios y, mantener la competencia en el mercado (Fabra & Toro, 2005).

Con esta regulación el principal objetivo del Estado era crear un mercado libre en el que los precios se determinan de forma natural mediante competición y no dependen de la imposición de poder del líder del monopolio, para con ello conseguir una mayor eficiencia de precios para los consumidores. El foco de la regulación estaba en conseguir un precio de las tarifas asequible para el consumidor y en evitar que las empresas se adquirieran entre ellas para no volver al monopolio natural inicial. Con las medidas aplicadas desde 1997 y fuente de la subestimación del coste que implicaban los cambios, se acumuló de forma masiva el déficit de tarifa. A la vista de estos hechos surgieron dudas sobre la constante actualización de la normativa a lo largo de los años en los que estuvo vigente la Ley 54/1997 y sobre la sostenibilidad económica del sector (Crampes & Fabra, 2005; López, 2022).

2.2.2. Ley 24/2013, de 26 de diciembre, del Sector Eléctrico²

La Ley 24/2013 derogó la Ley 54/1997 con la intención de integrar los diversos cambios y numerosos decretos publicados entre ambas leyes en un único instrumento más sencillo de interpretar y ejecutar. Esta ley también introduce una nueva forma de funcionamiento del mercado en cuanto a la determinación de la tarifa, que será detallado más adelante en esta investigación, y medidas urgentes para conseguir estabilidad financiera en el mercado eléctrico, entre las cuales se incluye la estimulación de la inversión en el sector para aumentar la competencia en las partes del mercado que funcionan en régimen de libertad (Díaz Mendoza et al., 2015; Pérez de Arévalo, 2021). Por lo que aunque la ley de 2013 no introduce grandes cambios, sí trata de remediar los fallos del ámbito económico y financiero del sistema anterior, especialmente el déficit de tarifa (Azabal, 2022).

2.3. Estructura y funcionamiento

Uno de los principales cambios que introdujo la ley de 1997, y que se mantuvo en la ley de 2013, es la división del mercado en cuatro partes, junto con la introducción de un

² BOE (2013): “Ley 24/2013, de 26 de diciembre, del Sector Eléctrico”. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2013-13645> [consulta: 02 de mayo de 2023]

nuevo agente, el operador. Esto se hizo porque las empresas que formaban el mercado hasta ese momento no solo presentaban una combinación diversa de estructuras de generación energética, sino que también estaban integradas verticalmente. La misma empresa formaba parte de cada una de las etapas por las que pasa la electricidad desde su producción hasta su llegada a los hogares de los consumidores finales (Agosti et al., 2007; Méndez & Santoyo, 2020). Después de la introducción de este cambio las empresas antiguas se separan en distintas sociedades para poder seguir realizando las mismas funciones, a diferencia de las que entran a partir de la regulación de 1997, que se enfocan en solo una de las partes del mercado y, por lo general, en un único tipo de generación de electricidad. Esta disparidad supone una barrera de entrada para los nuevos integrantes del mercado ya que están en desventaja a la hora de competir en precios y ofertas (Vives, 2006).

2.3.1. Agentes del mercado eléctrico:

- *Generación*

Se encargan de la producción las personas físicas o jurídicas que dedican su actividad a transformar los recursos en electricidad y verterla al mercado a cambio de un precio. Su principal función es la transformación de las materias primas en electricidad, además de la construcción, mantenimiento, y operación de las infraestructuras que se necesitan para llevar a cabo esa función (Pérez de Arévalo, 2021).

Esta parte del mercado ha evolucionado desde la Ley 54/1997 por la liberalización que ha sufrido y por la introducción a gran escala de las energías renovables en el esquema de generación de electricidad. Esto ocurrió especialmente a partir de 2007 que es cuando se publicó el Real Decreto 661/2007³ (RD en adelante). Dicho RD favorecía la creación e inversión en las tecnologías de generación de energía renovable por medio de primas generosas (Ciarreta et al., 2017; Díaz Mendoza et al., 2015). El problema principal de la electricidad que procede de fuentes de origen renovable es que su producción no es constante porque depende de factores externos que no son confiables al 100%. Por ello

³ BOE (2007): “Real Decreto 661/2007, de 25 de mayo, por el que se regula la actividad de producción de energía eléctrica en régimen especial.”. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2007-10556> [consulta: 02 de mayo de 2023]

no puede ser fuente única de generación a pesar de ser económicamente más asequible. Se sigue necesitando un porcentaje de producción de energía no renovable, cuya fuente de origen es más costosa pero gestionable, especialmente en momentos con tanta fluctuación en la demanda como en la actualidad (Arcos, 2019; López, 2022).

- *Transporte*

Es una de las partes del mercado cuya regulación no cambió de forma drástica en 1997, ya que se mantiene el régimen de monopolio natural bajo la compañía Red Eléctrica Española (en adelante REE). Además de transportar la electricidad a lo largo de líneas de alta y media tensión, debe asegurarse de que la pérdida a lo largo del recorrido hasta el suministro sea mínima. En esta función se mantuvo la competencia única de la compañía REE porque su liberalización no daba lugar a eficiencias, sino que podía perjudicar al consumidor final. Se consideró importante desde el principio que la red de transporte fuera igual de completa y eficiente para todos (Bonilla, 2017; Córdoba, 2012; García-Abril, 2022).

- *Distribución*

La distribución de la electricidad es la parte de la cadena que conecta la red de transporte con el consumidor final. Esta actividad se lleva a cabo por medio de varias compañías según el territorio, pero actúan en régimen de oligopolio regulado, pues todas tienen los mismos derechos y obligaciones entre ellas y con respecto a la REE (Córdoba, 2012; García-Abril, 2022; Méndez & Santoyo, 2020). Su función no solo es la de distribuir la electricidad a los puntos de consumo, también debe asegurar el suministro contratado por la comercializadora, mantener las instalaciones y controlar los contadores que miden la cantidad de electricidad que utiliza cada consumidor (Redondo, 2015).

- *Comercialización*

La fase de comercialización de la electricidad contempla la compra de esta a los productores y la venta al consumidor final mediante las transacciones necesarias para que se transporte y distribuya desde el punto de generación al de consumo final (Pérez de Arévalo, 2021). Hay dos tipos de comercializadora: las comercializadoras de referencia o de último recurso, que ofrecen una tarifa a un precio regulado, y las comercializadoras libres, cuya tarifa depende de una parte regulada y una libre basada en la energía consumida y las condiciones pactadas por el consumidor con la comercializadora. Las

funciones de las compañías comercializadoras incluyen captar clientes y gestionar sus contratos, además de controlar la cartera completa de consumidores y la estimación de su demanda (Córdoba, 2012; Méndez & Santoyo, 2020). Además las compañías comercializadoras deben pagar por adelantado ciertos peajes e impuestos por cada uno de sus clientes; estos se corresponden con una parte regulada que pagará el consumidor durante su contrato.

- *Operador*

En Europa hay una entidad formada por todos los operadores de los mercados eléctricos que forman la Unión Europea: este es el Mercado Interior de la Electricidad (López Domínguez, 2022). Actualmente y desde 2004 el mercado eléctrico de la península se integra completamente uniendo España y Portugal bajo un único sistema, el Mercado Ibérico de la Electricidad (en adelante MIBEL); este actúa en exclusividad en todo el territorio facilitando la cooperación y la estabilidad. Esta entidad forma parte y representa a ambos países en el Mercado Interior de la Electricidad (MIBEL, 2009).

De MIBEL dependen dos sociedades que son el Operador del Mercado Ibérico- Polo Español y el Operador del Mercado Ibérico- Polo Portugués (en adelante OMIE y OMIP). A pesar de que sus funciones son diferentes, las tres entidades actúan bajo los mismos principios: transparencia, objetividad e independencia del Estado (Pérez de Arévalo, 2021; Velázquez, 2018).

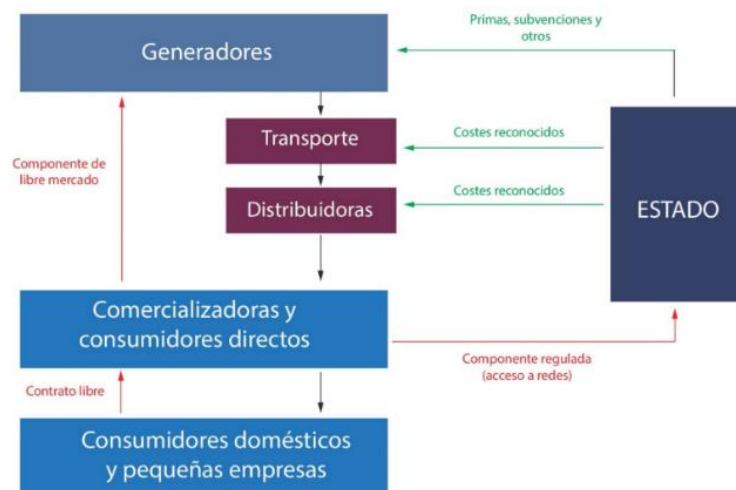
El OMIE se encarga de operar el mercado *spot* o al contado, regular, organizar y liquidar las compras y ventas del mercado español día a día de forma imparcial a este, además de facilitar la comunicación entre las distintas partes del mercado (Enseñat, 2022; Sánchez, 2022). El OMIP se encarga del mercado a futuro o a plazos en el que se negocian precios de la electricidad previos a la producción de esta desde dos días hasta tres años de antelación (Bonilla, 2017; Velázquez, 2018).

2.3.2. Funcionamiento:

En la figura 1 puede observarse de forma esquemática la relación que existe entre las distintas parte del sector energético, las flechas rojas y verdes se corresponden con flujos de dinero y las negras con la energía. El funcionamiento del sector de la energía y el sistema por el que se determinan los precios de este dependen de dos mercados: el

mayorista y el minorista. El mayorista es en el que participan las productoras de electricidad, las comercializadoras y los grandes compradores. En este mercado hay dos tipos de productos: por un lado, los contratos bilaterales entre dos partes que suponen un porcentaje minoritario de las negociaciones, y, por otro lado, el *pool* que es donde ocurren la mayoría de las operaciones. Las compañías generadoras vierten su flujo de electricidad producida al mercado *pool*, cuyo nombre hace referencia al significado de la palabra, que en español es piscina, porque es el lugar en que se almacena toda la energía creada por las empresas generadoras independientemente del tipo y del origen. Las comercializadoras y grandes compradores que tienen acceso al *pool* de energía compran las cantidades que necesitan en función de la demanda de los clientes (Azabal, 2022).

Figura 1: Funcionamiento del sistema en España



Fuente: Bonilla Martínez, 2017

El mercado minorista es el que se da entre las grandes comercializadoras y los consumidores finales. Tiene una parte libre, en la cual el consumidor puede elegir la compañía, la tarifa y el contrato según sus necesidades, en función de los MW de potencia y el precio de las distintas ofertas, y una parte regulada, llamada Precio Voluntario para el Pequeño Consumidor (en adelante PVPC), en la cual se adjudica la comercializadora al cliente final. La tarifa del PVPC está limitada a consumidores con una potencia máxima de 10 MW y su precio, que está regulado por el Estado, depende directamente de los cambios en el mercado *pool* (Sánchez, 2022).

La entidad que opera el mercado *pool* es el OMIE. Entre sus funciones principales se encuentra la de determinar los precios diarios de la electricidad según las curvas de oferta

y demanda. El OMIE recibe todas las ofertas de ambas partes para el día siguiente, las ordena y crea dos curvas una de electricidad ofertada y otra de la demandada, siendo el punto de cruce de ambas es el precio de venta de la energía. Este proceso es marginal, es decir el precio que se determina es común para todos los vendedores y compradores y lo determina la última unidad de energía necesaria para cubrir la demanda (Barrios, 2021). Cada hora habrá un precio fijo y estable para todos los tipos de energía que formen el *pool* de ese periodo porque la electricidad final es la misma independientemente de la fuente y el coste de su generación. Este tipo de determinación de precios beneficia a los ofertantes de energías más baratas y a los compradores dispuestos a pagar precios más altos (Enseñat, 2022).

- *Curva de oferta de las productoras*

Las empresas productoras determinan la curva de precios a la que se oferta la energía para cada hora. El precio no se calcula según los costes fijos y variables, sino que depende del coste de oportunidad; es decir, los costes que evita al producir la energía y venderla (detener la maquinaria, costes fijos independientes del funcionamiento, etc.) y los ingresos que deja de ganar por verter su energía a la red en ese momento (generar energía en una franja horaria más cara). Las compañías que se dedican a la generación de energía reciben los denominados ingresos por capacidad, que sirven para evitar que suban los precios excesivamente y para promover la inversión en nuevas tecnologías mediante una mayor rentabilización de esta parte del sector. Estos ingresos son aparte de los correspondientes a la venta de energía en el mercado *pool* y los contratos bilaterales (Córdoba, 2012).

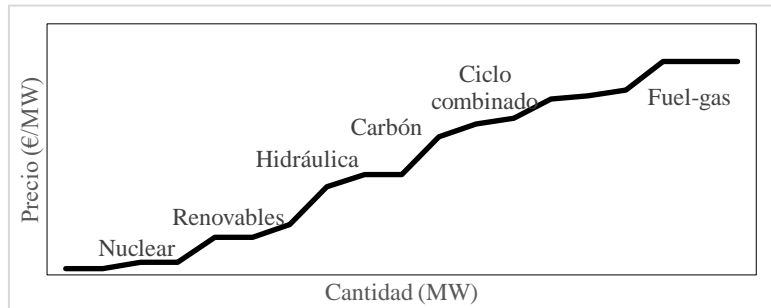
Los distintos tipos de tecnologías de generación de electricidad son: la generación nuclear, la de origen renovable y la térmica (Enseñat, 2022; Redondo, 2015).

- Generación nuclear: esta tecnología se define por la rigidez y la dificultad a la hora de modificar su capacidad de producción. Es por ello que las generadoras de energía nuclear ofertan su producción a precio cero, para asegurar que toda su producción se cubre en el mercado y que no tienen que detener la maquinaria. Por otro lado, la ventaja de este tipo de generación es que es constante, estable y, por lo tanto, supone una fuente segura para el mercado.

- Generación de origen renovable: la principal característica de este tipo es que utiliza fuentes inagotables como el agua, el sol y el viento. Además la producción de gases de efecto invernadero de este tipo de tecnologías es muy bajo o incluso nulo, lo que reduce el coste considerablemente y respeta el medio ambiente. El problema de la energía de origen renovable es que depende de la cantidad y constancia de las fuentes de origen, que por lo general son estacionarias y poco confiables. Las generadoras de este tipo de electricidad ofertan a precios muy bajos porque cómo dependen del flujo de la fuente de origen y no pueden almacenar la producción, prefieren venderlo a un precio menor antes que desperdiciarlo. La excepción son las presas hidráulicas que tienen más control sobre el momento de producción y, por ello, ofertan a precios un poco más altos.
- Generación térmica: las materias primas que utiliza este tipo de tecnología incluyen el carbón, el gas, los ciclos combinados y la cogeneración. Suponen el siguiente escalón de precios de la curva de oferta por el alto coste de los combustibles que utilizan, las tasas que deben pagar por la gran cantidad de gases de efecto invernadero que generan y por la flexibilidad de decisión que tienen las generadoras sobre el momento y la cantidad a producir. La mayor parte de la electricidad de este origen se importa de otros países, lo que también se refleja en sus altos precios.

El operador del mercado, OMIE, recibe las ofertas de todas las generadoras de electricidad y las ordena de más baratas a más caras según la cantidad ofertada para crear la curva de oferta del mercado (Figura 2). Las tecnologías que suelen definir el precio de mercado son la hidráulica y la térmica (carbón) según la demanda, cuanto más haya más alto será el precio por la necesidad de incluir en la curva las energías con precios más caros, que por lo general solo sirven para cubrir picos de necesidad (Azabal, 2022).

Figura 2: Curva de oferta de electricidad por tipo de origen y tecnología generadora

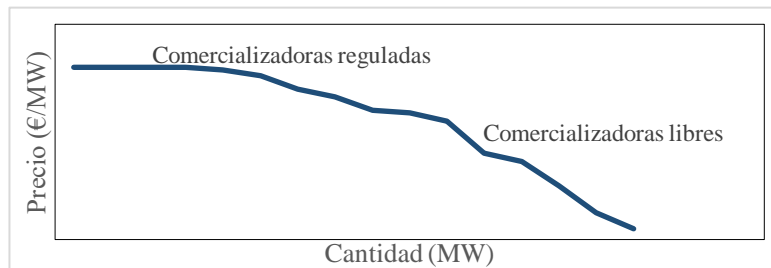


Fuente: elaboración propia

- Curva de demanda de las comercializadoras

La demanda depende de los consumidores finales que tienen tarifas y contratos con las comercializadoras que realizan las ofertas. En la figura 3 se puede observar que la primera parte de la curva la forman las ofertas de las comercializadoras reguladas (PVPC); estas están dispuestas a pagar precios más altos para asegurar que toda su demanda se cubre. El resto de la curva depende de las comercializadoras libres que tienen una mayor diversidad de precios y tarifas. Algunas deciden su demanda en función del precio final, estas son las que componen el final de la curva (Enseñat, 2022).

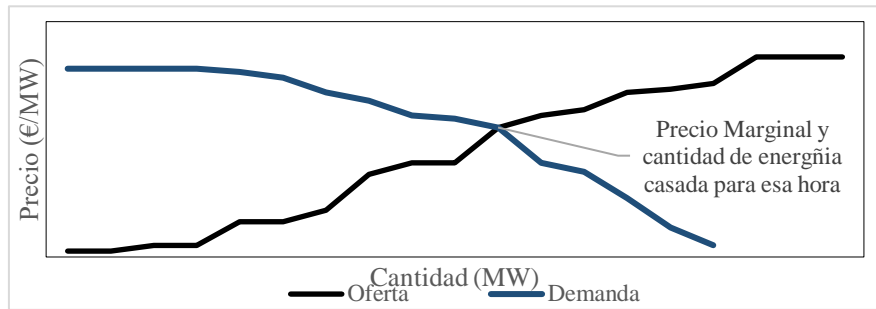
Figura 3: Curva de demanda de electricidad por tipo de comercializadora



Fuente: elaboración propia

El OMIE se encarga de ordenar las ofertas de demanda de mayor a menor en función del precio dispuesto a pagar por parte de la comercializadora. El punto de corte de ambas curvas es el que determina el precio marginal de la energía para cada hora. Como se puede observar en la figura 4, cuanto mayor cantidad de energía de origen renovable y nuclear se genere o menor demanda haya, más bajo será el precio de corte. En cambio, en momentos de alta demanda y fluctuaciones en las predicciones de esta se debe recurrir a energías más caras y, por lo tanto, el precio en el mercado *pool* se incrementa (López, 2022).

Figura 4: Curva demanda y oferta agregada



Fuente: elaboración propia

Las compañías que están integradas verticalmente forman parte de varias partes del mercado y las que lo están horizontalmente se benefician tanto del PVPC, como del mercado libre. Por lo que las que están completamente integradas tienen márgenes menos ajustados y más poder con respecto al resto. Por eso en la regulación se incluyen distintos incentivos para evitar que estas suban o bajen los precios según sus intereses estratégicos y no los del consumidor (Córdoba, 2012; López, 2022). Pero eso no ha evitado que a lo largo de los años se hayan dado numerosas guerras de precios. A excepción de esos momentos en los que han primado los intereses estratégicos de las compañías, el mercado funciona bajo colusión. Es decir, se busca llegar a un acuerdo sobre el precio justo para la electricidad, que es un bien necesario (Fabra & Toro, 2005).

2.3.3. Tasa de abandono

El negocio de las comercializadoras de electricidad está en crisis en la actualidad. La rotación de clientes ha aumentado considerablemente en los últimos años, al igual que la tasa de abandono. En el caso de los datos que se van a utilizar para el análisis de este trabajo de fin de grado, de los nuevos clientes que entran a formar parte de la cartera en 2021 un 68% abandonan antes de terminar el año. Las compañías se “roban” clientes las unas a las otras continuamente, lo que genera incertidumbre e inestabilidad en la curva de la demanda y en la rentabilidad económica a largo plazo de las propias comercializadoras. Por eso, una de las herramientas esenciales para las compañías comercializadoras de productos y servicios de mercados, en los que hay mucha competencia y el producto a vender es muy similar en todas ellas, es la Gestión de la Relación con los Clientes (*Customer Relationship Management*, CRM en adelante).

El CRM se enfoca en crear y mantener relaciones con los clientes que aporten beneficios mutuos para ambas partes y sean a largo plazo (Ngai et al. 2009). Por lo tanto, las

principales tareas del CRM son la atracción de nuevos clientes y la retención de los que ya forman parte de la cartera. La primera es la que más costosa para las compañías, por la inversión en publicidad y visibilidad que requiere y el bajo efecto final que tiene en la rentabilidad. La segunda es la más beneficiosa en cuanto a costes y beneficios, pero la más complicada en la actualidad por la alta rotación que hay en el mercado (Buckinx, 2005; Martínez, 2004).

Una de las técnicas que utiliza el CRM para evaluar la retención de clientes de la compañía es la tasa de abandono (*Customer Churn*⁴, en adelante *churn*). El *churn* mide el porcentaje de clientes que prescinde de los servicios de la compañía para contratar los de otra sobre el total de la cartera en un periodo y tiempo determinados (Chen et al., 2012; Hashmi et al., 2013). El objetivo de las predicciones de *churn* a futuro es identificar los posibles clientes que van a abandonar la compañía y las variables que más efecto tienen sobre esa decisión.

Se ha realizado este tipo de estudios de forma extensiva en mercados como el de las telecomunicaciones (Hung et al., 2006; Verbeke et al., 2012) y los servicios financieros (Glady, 2009) que, en condiciones de mercado, son parecidos a la comercialización de energía y la diferencia principal está en el propio producto, porque como se ha resaltado a lo largo de la investigación, la electricidad no es almacenable.

Los modelos de predicción tratan de saber que clientes de la cartera van a querer abandonar la compañía antes incluso de que estos lo sepan, basándose en la información de los que ya han abandonado anteriormente. Con este conocimiento las comercializadoras tienen un margen de maniobra para contactar al cliente antes de que este cancele el contrato y ofrecerle los cambios necesarios para que no abandone finalmente la compañía (Pribil & Polejova, 2017).

⁴ La palabra *churn* se utiliza como sinónimo de abandono, como referencia del abandono de los clientes en las compañías.

3. Metodología

3.1. Datos y variables

Los datos empleados en el análisis de esta investigación son reales, provienen de una de las empresas comercializadoras del país que por razones de privacidad y seguridad va a permanecer como fuente anónima. Para cumplir con esta directiva se han cambiado los valores de las variables *cups*⁵ y producto, pero se han mantenido sus características principales. El estudio se enfoca en una base de datos de una compañía particular y, por lo tanto, los modelos predictivos y conclusiones de la investigación son específicos de esta y no generalizables al resto de empresas que forman el mercado.

Se trata de la base de datos que forma la cartera de clientes de la empresa a fecha de octubre de 2022. Esta está formada por todos los clientes activos en ese momento y los que han abandonado entre enero de 2021 y noviembre de 2022, originalmente suponen un total de 140.674 registros en la base. Las variables que describen a cada cliente se resumen en la tabla 1. Se divide en variables originales, es decir, obtenidas directamente de la empresa y sin modificar y en variable creadas, que son aquellas que se han calculado a partir de otras variables o información externa para añadir información a la base de datos.

⁵ El cups se corresponde con las iniciales de Código Universal de Punto de Suministro, es un código único de identificación de cada punto de suministro de electricidad.

Tabla 1: Variables de la base de datos

	Tipo	Valores
Contrato Primero Fecha Inicio	Fecha	Julio 2009 - octubre 2022
Contrato Actual Fecha Inicio	Fecha	Julio 2009 - octubre 2022
Contrato Actual Fecha Fin	Fecha	Enero 2021 - diciembre 2030
Consumo Estimado (kWh)	Numérica continua	100 - 14.000
Cups	Cualitativa nominal	1 - 123.831
Producto	Cualitativa nominal	Fijo 1 - Fijo 4
		Variable 1 - Variable 6
Tarifa	Cualitativa nominal	Residencial
Gestor	Cualitativa nominal dicotómica	Gestor Interno/ Canal
Rotación	Cualitativa nominal dicotómica	Duplicado/ Único
Duración	Numérica discreta	30 - 5.127 días
Precio Medio OMIE	Numérica continua	Valor 0 para los no finalizados
		Precio medio para los finalizados
<i>Churn</i>	Cualitativa nominal dicotómica	0 <i>Churn</i> / 1 No <i>Churn</i>

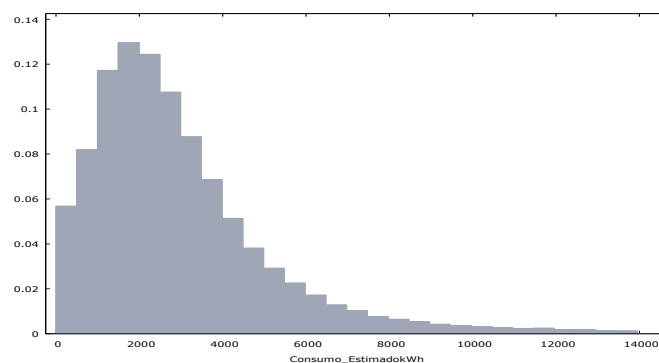
Fuente: Elaboración propia

3.1.1. Variables originales

- *Cups*: Valor único e irrepitible que identifica cada punto de conexión entre un hogar y la red. El valor se ha cambiado para proteger la privacidad de la empresa y los clientes, pero se mantiene la relación clave-valor, por lo que si un consumidor está duplicado es identificable mediante esta variable. Toma valores entre el 1 y el 123.831 lo que implica que hay clientes que se duplican, porque el número de clientes es inferior al número de filas total de la base.
- Contrato primero fecha de inicio: se trata de la fecha en la que el cliente inicia su contrato. Si un cliente abandona la empresa y vuelve, se crea una nueva fecha de inicio porque, aunque es el mismo cliente, es otro contrato distinto. Esto ocurre en todos aquellos *cups* que están duplicados y nos permite identificar clientes que han rotado en el mercado.

- Contrato actual fecha inicio: identifica la última fecha en la que ese cliente ha renovado su contrato, que en algunos casos coincide con la fecha del primer contrato.
- Contrato actual fecha fin: señala la fecha en la que el contrato termina. En los casos en los que el cliente aún no ha abandonado la compañía es la fecha en la que debe renovar el contrato y en el resto de los casos es la fecha en la que el cliente ha dejado la compañía.
- Consumo estimado: es la estimación media del consumo en kWh de cada cliente. Se han eliminado aquellos valores que se consideran *outliers* para el análisis que se está realizando y la categoría que se quiere analizar, que es residencial, esto incluye los consumos inferiores a 100kWh y superiores a 14000 kWh. En la Figura 5 puede observarse como se distribuyen los valores de la variable.

Figura 5: Distribución de frecuencias de la variable consumo estimado

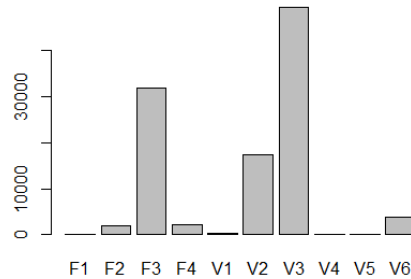


Fuente: Elaboración propia

- Tarifa: es una variable categórica que diferencia a los clientes en tres tipos residencial, pymes y medianas empresas e industria. Aunque esta investigación se va a enfocar solo en la categoría residencial que es el segmento en el que más rotación y abandono de clientes hay.
- Producto: es una variable categórica cuyos valores representan las distintas tarifas que ofrece la compañía, que son diez, cuatro de precio fijo y seis de precio variable. Los nombres reales de los productos que ofrece la compañía han sido sustituidos por otros en los que se identifica si el precio es de tipo fijo (F) o variable (V) y diferencia entre productos dentro de cada tipo. Debido al bajo número de clientes que forman algunas de las categorías, como se puede observar

en la Figura 6, se ha decidido agrupar bajo “otrosF” los clientes que tienen los productos F1, F2 y F4 y “otrosV” los que tienen los productos V1, V4 y V5.

Figura 6: Distribución de los valores originales de la variable Producto



Fuente: Elaboración propia

- Gestor: es una variable dicotómica que diferencia a los clientes según si han contratado el producto mediante un gestor interno de la compañía o un canal de comercialización externo, por lo que toma los siguientes valores: Gestor Interno o Canal.

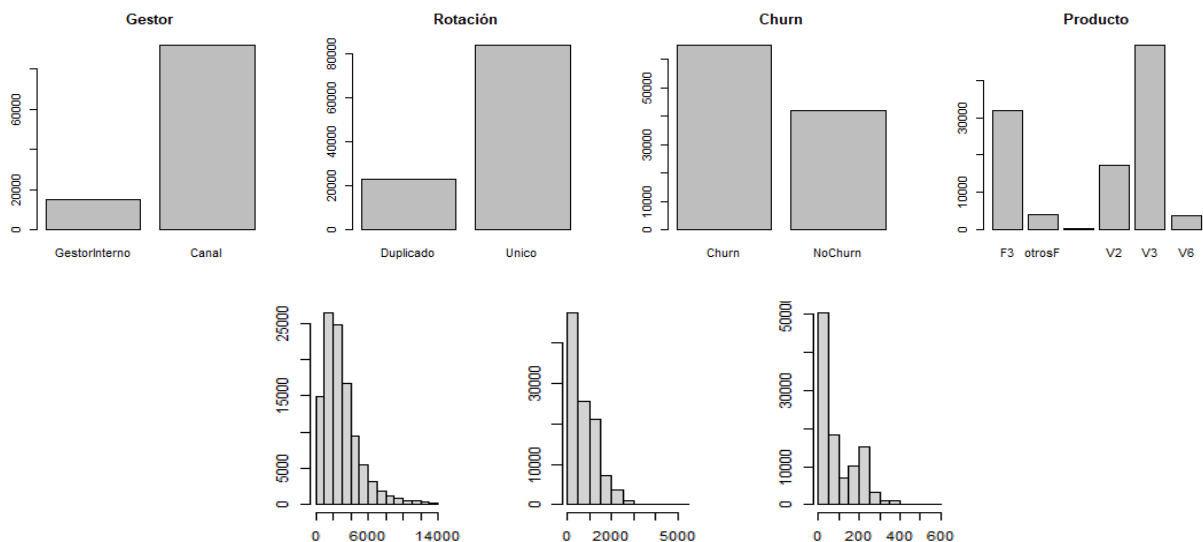
3.1.2. Variables creadas

- En el caso específico de esta base de datos y en relación con lo que se está estudiando que es el abandono y la duración de los clientes en la empresa, se ha decidido no eliminar los *cups* duplicados, porque no son una repetición de la información, si no que identifican aquellos clientes que ha rotado al menos una vez por el mercado, es decir han abandonado la compañía y han vuelto más tarde. Por lo que la primera variable creada es la de “rotación” que diferencia entre si el cliente ha dejado la compañía previamente o no, según si el *cups* está duplicado o no.
- La siguiente variable que se ha creado para el análisis es la de “duración” que mide la diferencia entre la variable “contrato actual fecha fin” y la de “contrato primero fecha inicio”. Para aquellos clientes que han abandonado la compañía, representa la duración real de su vida en la empresa y para los que no, mide la duración hasta la siguiente renovación de contrato. Se han eliminado aquellas filas cuya duración es inferior a treinta días, porque en muchos de los casos coincide con errores que se solucionan creando un nuevo contrato y no con abandonos reales.

- La variable “precio medio” es el precio medio de la electricidad correspondiente a dos meses antes de la fecha de abandono del cliente según los precios que ofrece el OMIE. En el caso de los consumidores que no han abandonado la compañía el valor de esta variable es cero. Debido a la lógica que toma esta variable será excluida del modelo de predicción del abandono, ya que sus valores toman la misma dicotomía que la variable a analizar: aquellos que abandonan tienen un valor de precio medio y los que no su valor es cero.
- La variable “abandono” es dicotómica, toma valor cero si la fecha fin del contrato es anterior a la fecha de descarga y, por lo tanto, representa a aquellos casos que han abandonado la compañía y toma valor uno para los que tienen una fecha fin posterior a octubre de 2022, es decir, lo que no han abandonado la compañía.

Se puede observar en la figura 7 la representación gráfica de la distribución de los valores de cada una de las variables que se van a utilizar para realizar los modelos. Las variables de fecha de inicio y fin y de valor de cups se han excluido porque solo se han utilizado para crear otras variables, pero no van a formar parte de los propios modelos. En el *dataset* hay un total de 64.773 clientes que abandonan (61%) y 41.891 que no abandonan (39%) y de los que abandonan un 35% (22.810 clientes) ya lo habían hecho previamente.

Figura 7: Representación de las variables



Fuente: Elaboración propia

3.2. Análisis

En este trabajo de fin de grado se quiere analizar las variables creadas “abandono” y “duración”.

El primer enfoque que se utiliza es una predicción del abandono de los clientes mediante una regresión logística múltiple; se utiliza este método por el formato dicotómico de las variables a predecir. También se decide utilizar un segundo método para el análisis de la duración por la lógica de la propia variable, ya que la duración solo es exacta para aquellos casos en los que el cliente ha abandonado. Para el resto se tiene el valor de la duración actual hasta la próxima renovación del contrato prevista, que no tiene por qué corresponderse con el momento real en que el cliente abandone la compañía. Para tener en cuenta esta información se realiza un *survival analysis*.

3.2.1. Predicción de la tasa de abandono: Regresión

Diversos autores, algunos de ellos citados en este capítulo, han realizado revisiones bibliográficas sobre la predicción del *churn* en el mercado eléctrico y han afirmado que hay poca información y pocos estudios enfocados en este tema. La mayoría utilizan como referencia investigaciones con el mismo enfoque aplicadas a otros mercados, como el de las telecomunicaciones, que funciona de una forma similar.

Según Neslin et al. (2006) un modelo que se utiliza frecuentemente en la predicción del *churn* es la regresión múltiple logística. La regresión se considera de los métodos más ajustados en la predicción de la probabilidad de abandono de los clientes y, específicamente, se prefiere la logística por la dicotomía de la variable a predecir; en cambio la regresión lineal se ajusta mejor a variables numéricas (Huang et al., 2012; Lazarov & Capota, 2007). Aunque la decisión sobre el mejor modelo de predicción se toma en función de la bondad de los resultados y la robustez de las predicciones, también hay otras razones en las que basarse, como pueden ser el tiempo del que se dispone para probar y correr modelos y los conocimientos del analista que realice la investigación. Es por estas razones también que el modelo preferido para empezar a predecir el *churn* es la regresión logística, porque es sencilla, rápida y, por lo general, efectiva en cuanto a los resultados (Marttinen, 2021).

Modelo

La regresión logística es un modelo que relaciona una variable dependiente (Y) categórica con una o más variables independientes (X1, X2,... Xn) categóricas o numéricas. Calcula la probabilidad de que ocurra Y, es decir que la variable independiente dicotómica sea 1, según los valores que tomen el resto de las variables X. En este caso, Y se corresponde con la variable *churn* que es un factor que puede tomar dos valores y el modelo clasifica cada observación según estos, *churn* o *no churn*, Para llevar a cabo esta función la regresión se entrena con el porcentaje de los datos establecido para ello y prueba su capacidad de predicción con respecto a si el cliente abandona (*churn*) o no (*no churn*) con el resto. Además este proceso se ha repetido varias veces mediante validación cruzada.

El modelo se ha llevado a cabo en RStudio que es un *software* libre con el que se pueden realizar estudios estadísticos y representaciones gráficas de estos. En el desarrollo del modelo de regresión se han utilizado distintos métodos a la hora de buscar el mejor resultado. Por un lado, se ha utilizado la división de datos en entrenamiento y prueba según distintos porcentajes, específicamente se ha realizado la división 80/20, 70/30 y 60/40. Esto significa que, por ejemplo, en el primer caso, el 80% de los datos han entrenado el modelo de predicción y el 20% restante se ha utilizado para evaluar su calidad. También se ha realizado validación cruzada de 5 y 10 repeticiones para cada escenario y todo ello se ha aplicado a un total de 10 combinaciones diferentes de las variables, por lo que en total se han realizado 60 modelos. Las validaciones cruzadas arrojan valores similares en todos los casos y la división de datos en 80% entrenamiento y 20% prueba es la que mejor resultado tiene. El código del modelo con el cual se obtienen los mejores resultados está en el anexo 1. De todos los modelos comparados, el que mejor predice es:

$$\text{Churn} = \text{Gestor} + \text{Rotación} + \text{Duración} + \text{Consumo} + \text{Producto} + \text{Gestor} \times \text{Rotación}$$

En este caso, se ha relacionado la variable gestor y rotación, por el vínculo que existe actualmente en el mercado del uso por parte de las comercializadoras de canales de venta externos o subcontratados y la tendencia de los clientes a rotar por el sector.

- *Resultados*

En un modelo de regresión logística se obtienen una serie de coeficientes en relación con cada variable independiente. Con estos valores se crea la función de regresión mediante la cual se estima el valor logarítmico de la probabilidad (*logit* (p)) de que ocurra la variable a predecir, en este caso el valor *churn*, que se corresponde con los clientes que sí abandonan. Este valor no es directamente interpretable: se le tiene que aplicar la función logística que es exponencial y convierte el valor *logit*(p) en una probabilidad que oscila entre 0 y 1 (Kleinbaum & Klein, 1996).

La función de regresión (*logit* (p)) del modelo final se corresponde con:

$$\begin{aligned} \text{Churn} = & 2,15 + 0,85 \times \text{Gestor (canal)} + 0,42 \times \text{Rotación (duplicado)} - 8,28 \times \text{Duración} + 2,62 \times \\ & \text{Consumo} - 2,14 \times \text{Producto (otrosF)} - 0,86 \times \text{Producto (otrosV)} - 1,96 \times \text{Producto (V2)} - 2,07 \times \\ & \text{Producto(V3)} - 2,45 \times \text{Producto (V6)} - 0,73 \times \text{Gestor (canal)} \times \text{Rotación (único)} \end{aligned}$$

Por lo general, el signo de los coeficientes representa el tipo de efecto y el valor indica el peso que tiene cada variable en la decisión de abandono de los clientes. Si es positivo, aumenta la probabilidad de que ocurra, y si es negativo, la disminuye; y este efecto será en mayor o menor medida según si el valor es más alto o tiende a cero. Pero hay distintas formas de explicar el valor de los coeficientes según las propiedades de la variable independiente a la que acompañan (Hosmer Jr et al., 2013; Pampel, 2020; Peláez, 2016):

- En el caso de las variables numéricas, como la duración y el consumo, los coeficientes se interpretan como el cambio proporcional en las oportunidades de abandono debido a una variación unitaria en la variable independiente correspondiente, manteniendo el resto constantes. Por ejemplo, si el coeficiente para la variable duración es -8,28 significa que un incremento de una unidad en la duración se asocia con una disminución proporcional de $e^{-8,28}$ en la probabilidad de que el cliente abandone. En otras palabras, a medida que la duración aumenta, el abandono tiende a reducirse.
- En cambio, en las variables categóricas lo que se hace es crear variables *dummy*. Se toma una de las posibles categorías como referencia y las demás se representan como nuevas variables binarias en la fórmula. Esto ocurre con todo el resto de variables, como por ejemplo la de producto, cuyo valor de referencia es F3. Cuando se interpreta el coeficiente asociado a una de las variables *dummy*, su signo (positivo o negativo) indica como se compara la categoría correspondiente

con la de referencia en términos de la probabilidad de abandono. Un coeficiente positivo significa que, en comparación con la categoría de referencia, la representada por la variable *dummy* tiene mayores posibilidades de abandono, y un coeficiente negativo tiene el significado opuesto. Por ejemplo, en este caso los coeficientes que acompañan a la variable producto en todas sus categorías son negativos, por lo que cualquiera de los productos tiene una menor probabilidad de abandono que el F3.

La calidad de la regresión se mide mediante la sensibilidad, la especificidad y la precisión. Los posibles resultados de las predicciones y sus significados pueden resumirse en la tabla 2 y la tabla 3, que se corresponde con una matriz de confusión:

Tabla 2: Valores de una matriz de confusión

Valor	Definición	Valor predicho	Valor real
A (verdadero positivo)	Indica el número de observaciones que pertenece a la clase que se predice (<i>churn</i>) y que el modelo ha clasificado correctamente en esa clase.	<i>churn</i>	<i>churn</i>
B (falso positivo)	Son aquellos valores que en la realidad no pertenecen a la clase que se está prediciendo, pero que el modelo ha clasificado incorrectamente en ese grupo.	<i>churn</i>	no <i>churn</i>
C (falso negativo):	Es la cantidad de observaciones que en la realidad si pertenecen al grupo <i>churn</i> , pero que el modelo clasifica incorrectamente como no pertenecientes.	no <i>churn</i>	<i>churn</i>
D (verdadero negativo)	Indica el número de observaciones que no forman parte de la clase a predecir y el modelo clasifica correctamente como no pertenecientes a esta.	no <i>churn</i>	no <i>churn</i>

Fuente: elaboración propia (a partir de Lazarov & Capota, 2007; Vafeiadis et al., 2015)

La sensibilidad se refiere al número de predicciones positivas (clientes que abandonan) correctas entre el total de predicciones positivas ($A/A+B$). La especificidad se refiere al número de predicciones negativas correctas (clientes que no abandonan) entre el total de los negativos ($D/(C+D)$). La precisión se calcula mediante la división de las predicciones correctas (positivos y negativos) entre el total de predicciones ($(A+D)/\text{Total}$) (Lazarov & Capota, 2007; Vafeiadis et al., 2015).

Finalmente, mediante la métrica *F-score* se mide el rendimiento general del modelo teniendo en cuenta tanto la capacidad de identificar correctamente los verdaderos

positivos, como la capacidad de evitar los falsos positivos y falsos negativos. El *F-score* se calcula de la siguiente forma:

$$F\text{-score} = 2 \times (\text{especificidad} \times \text{sensibilidad}) / (\text{especificidad} + \text{sensibilidad})$$

Tabla 3: Matriz de confusión

	<i>Churn</i> (Real)	No <i>Churn</i> (Real)
<i>Churn</i> (Predicción)	Positivo (A)	Falso positivo (B)
No <i>Churn</i> (Predicción)	Falso negativo (C)	Negativo (D)

Fuente: elaboración propia.

El rendimiento de este modelo para las distintas particiones de datos se puede ver en la tabla 4. Predice correctamente el 73,9% de los casos. A pesar de ser mejor que el azar, tiene un porcentaje de error demasiado alto para lo que se considera un buen intervalo de confianza en modelos de predicción. El principal problema que se encuentra la función es la disparidad que existe en los valores de la variable predictora: como se ha visto anteriormente el 61% de los clientes abandonan la compañía, mientras que el 39% no. Por lo general, en este tipo de predicciones es mejor que la clasificación esté equilibrada para que el nivel de error sea más bajo, pues en las métricas de evaluación del modelo se tiene en cuenta la capacidad de identificar correctamente tanto los valores positivos, como los negativos (Chen et al., 2004). De hecho, este problema puede observarse en la tabla 4 en la diferencia entre los valores que toman la sensibilidad que se refiere a la exactitud con la que el modelo mide los casos positivos que son la mayoría (61%) y la especificidad que, por lo contrario, representa la capacidad de predicción de casos negativos (39%) que tiene el modelo

Tabla 4: Rendimiento de la regresión logística para predecir el *churn*

	Partición 80/20	Partición 70/30	Partición 60/40
Precisión	73,9%	73,55%	73,46%
Sensibilidad	84,2%	84,12%	83,91%
Especificidad	57,9%	57,19%	57,30%
<i>F-score</i>	68,6%	68,1%	68,1%

Fuente: elaboración propia

3.2.2. Predicción del abandono de los clientes: *Survival analysis*

La regresión logística nos permite predecir con un alto grado de acierto los casos que abandonan, sin embargo, tiene un bajo rendimiento con respecto a la identificación de los casos que no lo hacen. Es por esto y por el alto peso que tiene según la regresión la variable duración en el abandono de los clientes, que se ha decidido realizar un segundo análisis que permite modelar la duración o tiempo hasta que ocurre un evento en particular, en este caso el *churn*.

- *Modelo*

El *survival analysis* es útil para realizar estudios en los que se pretende estudiar el tiempo entre dos eventos, el inicio y el fracaso. Se utiliza el término fracaso para denominar el segundo evento porque suele estar relacionado con un hecho que impacta de forma negativa. En este caso, el inicio se corresponde con el momento en que un nuevo cliente se une a la compañía y el fracaso con el momento en que abandona. (Kleinbaum & Klein, 1996). El uso de este análisis en este ámbito resulta novedoso, en la revisión bibliográfica realizada se ha encontrado poca información y, por lo general, ha sido con relación a otros temas, sobre todo aplicado a ramas de la salud. Aún así se ha considerado muy apropiado adaptarlo al análisis del abandono de los clientes en el sector energético.

La ventaja de este análisis es que tiene en cuenta la censura, que es la falta de información completa en una parte de los datos sobre el momento en que ocurre alguno de los eventos, ya sea porque no se tiene o porque aún no ha ocurrido (Lu, 2002). En la base utilizada para realizar este trabajo de fin de grado los datos censurados son aquellos que no han

abandonado la compañía. Para estos clientes se tiene un valor de duración calculado en base a los días que lleva en la compañía. Debido a que se tiene la información sobre el momento en que ocurre el primer suceso, pero no el segundo, estos datos están censurados por la derecha. Para identificarlos en el modelo se utiliza una nueva variable denominada censor, creada a partir del *churn*, que toma valor 1 en los fracasos (*churn*) y valor 0 en los censurados (*no churn*). Esta ventaja permite aprovechar la información de todos los clientes hasta el momento en que dejó de actualizarse la base de datos.

Kleinbaum & Klein (1996) lo explican de forma sencilla en su libro *Survival analysis: a self-learning text*. Este tipo de análisis se basa en dos funciones principalmente: la de supervivencia ($S(t)$), que muestra cómo cambia la probabilidad de que un cliente no abandone a medida que transcurre el tiempo, y la de riesgo ($h(t)$), que mide para aquellos clientes que aún no han abandonado la probabilidad de que ocurra el evento en un momento determinado futuro. Para realizar el *survival analysis*, se utiliza un modelo estadístico llamado regresión de Cox que permite examinar cómo influyen las diferentes variables en el riesgo de ocurrencia del evento de fracaso. Este modelo estima los coeficientes asociados a cada variable y proporciona las curvas de supervivencia ajustadas (George et al., 2014; Jenkins, 2005; Lu & Park, 2003).

Estas curvas muestran cómo cambia la probabilidad de que no ocurra el suceso a lo largo del tiempo para los diferentes valores de las variables y se calcula a partir del método de Kaplan-Meier. Este estimador se basa en el concepto de riesgo acumulado en cada punto de tiempo. Lo que hace es:

- Ordenar los datos de supervivencia
- Contar los sujetos en riesgo en cada momento
- Calcular la probabilidad de supervivencia multiplicando las probabilidades de momentos anteriores por la proporción de sujetos que sobreviven en cada momento, en el instante cero la probabilidad es 1
- Estimar las varianzas
- Con los valores obtenidos, representar en una gráfica la curva de supervivencia.

La mediana de la duración en la gráfica coincide con el valor de la X para una Y del 50% (Barranco, 2017).

- *Resultados*

Para realizar el *survival analysis* se ha utilizado Gretl, que es un *software* libre mediante el cual se pueden realizar tanto análisis econométricos, como estadísticos (Pérez, 2016). Para realizar el modelo se definen: la variable dependiente, en este caso la duración; la variable de censura, llamada censor en la base de datos; los regresores, que son el resto de variables a estudiar y utilizar para entrenar el modelo y el tipo de distribución. Debido a la poca información que hay publicada actualmente sobre el *survival analysis* aplicado al abandono de los clientes en el mercado energético, la decisión con respecto al tipo de distribución se ha tomado según los resultados obtenidos. Otra razón por la que se ha utilizado la distribución log-normal, es por la capacidad que tiene de ajustarse a procesos de crecimiento acumulativo, en este caso los eventos que han ocurrido anteriormente aumentan directamente la probabilidad de ocurrencia de eventos futuros (Kleinbaum & Klein, 1996).

La función de distribución de la duración estimada es:

$$X = e^{\mu + \sigma x Z}$$

En esta función hay una parte que es constante, formada por la multiplicación de sigma ($\sigma = 0,89$), que es un valor obtenido en el modelo, y Z que es el valor de la distribución normal. Y una parte que varía según los valores que toman cada una de las variables y da lugar a la curva: esta es μ que representa la ecuación formada por los coeficientes resultado del modelo.

$$\mu = 8,16 - 0,31 \times \text{Gestor (canal)} - 0,71 \times \text{Rotación (duplicado)} - 1,71 \text{ e-}05 \times \text{Consumo} - 0,29 \times \text{Producto (otrosF)} + 1,21 \times \text{Producto (otrosV)} + 0,51 \times \text{Producto (V2)} - 0,42 \times \text{Producto (V3)} + 0,45 \times \text{Producto (V6)} - 0,05 \times \text{Precio}$$

En este caso lo que se está prediciendo es la mediana de la duración según los valores que toman las variables. La explicación del valor de los coeficientes y el funcionamiento de los distintos tipos de variables es el mismo que en la regresión logística, lo que cambia es el valor que se está estimando y su significado. Las variables cuyo valor del coeficiente es positivo, aumentan la mediana de la duración, y cuando este es negativo disminuye. Por ejemplo, que el gestor sea un canal de venta externo disminuye el valor de la mediana y, por lo tanto, se estima que la duración del cliente en la compañía va a ser menor.

Si se comparan los valores obtenidos en ambos modelos se puede observar que muchos de ellos tienen un efecto similar. La relación entre ambos modelos es que cuando hay más probabilidad de abandono, la mediana de la duración tiende a ser menor. Por lo tanto, tiene sentido que aquellas variables cuyo coeficiente sea positivo en la regresión, el mismo sea negativo en el *survival analysis*.

En la tabla 5 se pueden observar los distintos coeficientes y sus implicaciones sobre las variables que estiman.

Tabla 5: Coeficientes de las variables y significado

Variable	Coeficiente en la regresión	Efecto sobre el <i>churn</i>	Coeficiente en el <i>survival analysis</i>	Efecto sobre la duración
Gestor (canal)	0,85	Aumenta	- 0,31	Reduce
Rotación (duplicado)	0,42	Aumenta	- 0,71	Reduce
Consumo	2,62	Aumenta	- 1,71 e-05	Reduce
Producto(otrosF)	- 2,14	Reduce	- 0,29	Reduce
Producto (otrosV)	- 0,86	Reduce	1,21	Aumenta
Producto (V2)	- 1,96	Reduce	0,51	Aumenta
Producto (V3)	- 2,07	Reduce	- 0,42	Reduce
Producto (V6)	- 2,45	Reduce	0,45	Aumenta
Duración	- 8,28	Reduce	Variable estimada	No aplica
Precio	Variable excluida	No aplica	- 0,05	Reduce

Fuente: elaboración propia

Las variables precio y duración no son comparables entre modelos porque no tienen valor de coeficiente en ambos, el precio se ha excluido de la regresión logística y la duración es la variable que se está prediciendo en el *survival analysis*. El tipo de gestor, la rotación de los clientes, el consumo y algunos productos (otrosV, V2 y V6) son las variables cuyo efecto se mantiene en ambos modelos. Por ejemplo, si la variable gestor toma el valor canal, la probabilidad de abandono aumenta y la mediana de la duración disminuye. Esta

relación entre modelos es importante debido a la falta de investigación e información que hay sobre este tema.

4. Conclusiones

La rotación de clientes en el mercado energético está aumentando especialmente en los últimos años. Este hecho es un problema por la necesidad que tiene este sector en específico de conocer la demanda futura en el corto plazo. Cada comercializadora de electricidad debe predecir la demanda diaria de su cartera de clientes, tarea que se ve afectada por el aumento en la rotación de estos de unas comercializadoras a otras.

Una posible solución a este problema es la creación de un modelo capaz de responder a las preguntas expuestas anteriormente: ¿Qué variables afectan más a la decisión de los clientes de dejar la compañía? ¿Qué variables determinan en mayor medida la duración del contrato de un cliente? Para así poder tratar de predecir las fluctuaciones de la cartera y tomar decisiones con respecto a ello que traten de minimizar el impacto e incluso evitar la situación.

La regresión logística múltiple no ha resultado eficiente en la predicción por la presencia de una mayoría de clientes que toman el valor *churn* (clientes que han abandonado la compañía) en comparación con los que toman el valor contrario de la variable a predecir, es decir los que no abandonan. Pero los valores de los coeficientes obtenidos en el modelo sirven para responder a la primera pregunta. Por ejemplo, la duración es la variable que más peso tiene en la decisión de abandono de los clientes, el valor obtenido (-8.28) implica que, a mayor duración, menos probabilidad de que el cliente abandone. Por otro lado, el consumo también tiene un peso considerable (2,62), pero en este caso el efecto que tiene es el contrario: cuanto más consumo, más probabilidad de que el cliente abandone. También tiene el mismo efecto que el gestor sea un canal externo (0,85) y que el cliente haya rotado previamente al menos una vez (0,42), ya que en ambos casos se estima un aumento de la probabilidad de abandono.

La segunda pregunta planteada en este trabajo de fin de grado hace referencia a la variable duración, para la cual se ha realizado un *survival analysis*. En este se observa que el efecto de algunas de las variables sobre la duración de los clientes (*survival analysis*) es el contrario al que obtienen en la predicción del abandono (regresión logística), esto puede

verse en los valores de los coeficientes. Por ejemplo, en el *survival analysis* que el gestor sea un canal externo (-0,31) y que el cliente haya rotado previamente (-0,71) reducen la mediana de la duración porque el valor de sus coeficientes es negativo, mientras que, como se ha señalado anteriormente, estos tomaban valor positivo en el primer modelo y, por lo tanto, aumentaban la probabilidad de abandono. Lo mismo ocurre con el consumo que, aunque en comparación con su resultado en la regresión ha perdido efecto, sí ocurre que este es el opuesto (-1,71e-05) en el segundo modelo.

Por lo que, la principal conclusión a la que se puede llegar es que, en la mayoría de las variables, ambos modelos están correlacionados inversamente. Es decir, aquellas variables que aumentan la probabilidad de que el cliente abandone, también tienden a disminuir la mediana de la duración. Esta relación, como se ha descrito anteriormente, se ve reflejada en el coeficiente de la variable duración en el modelo de regresión: este es alto y negativo, por lo que cada aumento unitario en esta variable reduce la probabilidad de que el cliente abandone. Es decir, la relación entre la duración y el abandono es inversa y, por lo tanto, aquellas variables que afectan positivamente a una de las dos variables lo harán de forma negativa a la otra y viceversa.

Una posible línea futura de investigación es la de adaptar estas dos opciones de modelos a otras compañías y tratar de analizar el sector en su conjunto, en vez de centrarse en una única empresa. El *survival analysis* no se ha aplicado en muchos campos para los que podría ser útil.

5. Bibliografía

- Agosti, L., Padilla, A. J., & Requejo, A. (2007). El mercado de generación eléctrica en España: estructura, funcionamiento y resultados. *Economía industrial*, 364, 21-37.
- Arcos Vargas, Á. (2019). Cambios contextuales en el sector eléctrico. Perspectivas del Consumidor, del Regulador y de la Sociedad.
- Azabal Encinas, A. (2022). El mercado eléctrico español.
- Barranco, M. R. (2017). Análisis de supervivencia: el estimador de Kaplan-Meier.
- Barrios, A. M. (15 de febrero de 2021). ¿Por qué el mercado eléctrico es marginalista? *El Periódico de La Energía*. <https://elperiodicodelaenergia.com/por-que-el-mercado-electrico-es-marginalista/>
- Bianco, V., Driha, O. M., & Sevilla-Jiménez, M. (2019). Effects of renewables deployment in the Spanish electricity generation sector. *Utilities Policy*, 56, 72-81.
- Blanco, O. A. (2011). ¿Es competitivo el mercado eléctrico español? Indicadores de abuso de poder de mercado y aplicación al caso de España. *Estudios de economía aplicada*, 29(2), 627-654.
- BOE (1997): “Ley 54/1997, de 27 de noviembre, del Sector Eléctrico”. Disponible en: <https://www.boe.es/buscar/doc.php?id=BOE-A-1997-25340> [consulta: 15 de febrero de 2022].
- BOE (2013): “Ley 24/2013, de 26 de diciembre, del Sector Eléctrico”. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2013-13645> [con-sulta: 15 de febrero de 2022]
- Böhringer, C. (2003). The Kyoto protocol: a review and perspectives. *Oxford Review of Economic Policy*, 19(3), 451-466.
- Bonilla Martínez, E. (2017). El mercado eléctrico español: estructura y formación de precios.
- Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European journal of operational research*, 164(1), 252-268.

- Burgos-Payán, M., Roldán-Fernández, J. M., Trigo-García, Á. L., Bermúdez-Ríos, J. M., & Riquelme-Santos, J. M. (2013). Costs and benefits of the renewable production of electricity in Spain. *Energy Policy*, 56, 259-270.
- Çelik, O., & Osmanoglu, U. O. (2019). Comparing to techniques used in customer churn analysis. *Journal of Multidisciplinary Developments*, 4(1), 30-38.
- Ciarreta, A., Espinosa, M. P., & Pizarro-Irizar, C. (2017). Has renewable energy induced competitive behavior in the Spanish electricity market? *Energy Policy*, 104, 171-182.
- Ciarreta, A., Pizarro-Irizar, C., & Zarraga, A. (2020). Renewable energy regulation and structural breaks: An empirical analysis of Spanish electricity price volatility. *Energy Economics*, 88, 104749.
- Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. University of California, Berkeley, 110(1-12), 24.
- Chen, Z. Y., Fan, Z. P., & Sun, M. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of operational research*, 223(2), 461-472.
- Córdoba Zatarain, M. (2012). Las empresas comercializadoras de electricidad en España.
- Crampe, C., & Fabra, N. (2005). The Spanish electricity industry: Plus ça change. *The Energy Journal*, 26(Special Issue).
- De Espana, R. E. (2012). The Spanish electricity system. Report. Available online: https://www.ree.es/sites/default/files/publication/2022/04/downloadable/avance_ISE_2021_EN.pdf.
- Díaz Mendoza, A.C.; Larrea Basterra, M.; Álvarez Pelegry, E.; Mosácula Atienza, C. De la liberalización (Ley 54/1997) a la reforma (Ley 24/2013) del sector eléctrico español. *Cuad. Orkestra* 2015, 10, 1–104
- Edmond, C. (2022, marzo 22). ¿Cuánta energía importa la UE de Rusia? *Foro Económico Mundial*. <https://es.weforum.org/agenda/2022/03/cuanta-energia-importa-la-ue-de-rusia/>
- Enseñat Saavedra, I. (2022). El mercado eléctrico mayorista español.
- Escribano, G., & Lázaro, L. (2020). Energía, clima y coronavirus. *Real Instituto Elcano*, 27.
- Fabra, N., & Toro, J. (2005). Price wars and collusion in the Spanish electricity market. *International Journal of Industrial Organization*, 23(3-4), 155-181.

- Fargas, F. A. (2016). Regulación y gestión de los recursos eléctricos durante el primer siglo de la industria eléctrica. *Revista de historia industrial*, 25(61), 111-139.
- García-Abril Curto, L. F. (2022). Funcionamiento del mercado eléctrico español. Una comparativa con el resto de Europa.
- García, P. (2022). Grupo Audax, bajo el que se integra Audax Renovables, multado con 9,2 millones de euros. Material Eléctrico - CdeComunicacion.es. <https://material-electrico.cdecomunicacion.es/noticias/sectoriales/55030/grupo-audax-multado-con-mas-nueve-millones-de-euros>
- Garrido-Perez, J. M., Barriopedro, D., García-Herrera, R., & Ordóñez, C. (2021). Impact of climate change on Spanish electricity demand. *Climatic Change*, 165, 1-18.
- Garrués-Irurzun, J. (2022). The Regulation of the Spanish Electricity Sector: A Long Way to Convergence. *Energies*, 15(19), 6931.
- George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of nuclear cardiology*, 21, 686-694.
- Gill-de-Albornoz, B., & Illueca, M. (2005). Earnings management under price regulation: Empirical evidence from the Spanish electricity industry. *Energy Economics*, 27(2), 279-304.
- Glady, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197(1), 402-411.
- González Hernández, C. (2022). Derecho de la energía: aproximación a la regulación eléctrica en España.
- Hashmi, N., Butt, N. A., & Iqbal, M. (2013). Customer churn prediction in telecommunication a decade review and classification. *International Journal of Computer Science Issues (IJCSI)*, 10(5), 271.
- Historia de la electricidad en España*. (2014, agosto 12). *Energía y Sociedad*. <https://www.energiaysociedad.es/manual-de-la-energia/1-2-historia-de-la-electricidad-en-espana/>
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

- Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1), 1414-1425.
- Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3), 515-524.
- Jenkins, S. P. (2005). Survival analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK, 42, 54-56.
- Jimeno, M. D. R. F., & Cebrián, M. S. (2015). El mercado eléctrico en España: La convivencia de un monopolio natural y el libre mercado. *Revista europea de derechos fundamentales*, (25), 257-297.
- KhakAbi, S., Gholamian, M. R., & Namvar, M. (2010). Data mining applications in customer churn management. In 2010 international conference on intelligent systems, modelling and simulation (pp. 220-225). IEEE.
- Kleinbaum, D. G., & Klein, M. (1996). Survival analysis a self-learning text. Springer.
- Lazarov, V., & Capota, M. (2007). Churn prediction. *Bus. Anal. Course. TUM Comput. Sci*, 33, 34.
- Lezaun, M. (2017). ¡Me han hecho slamming! *Escritura pública*, (103), 58-60.
- Llorente Gallardo, J. A. (2016). Análisis del mercado eléctrico en España.
- López Domínguez, M. I. (2022). El mercado eléctrico español. Evolución del precio de la electricidad en un entorno de crisis energética.
- López Pérez, V. (2022). Análisis de la rentabilidad de la comercialización de energía eléctrica en España.
- Lu, J. (2002). Predicting customer churn in the telecommunications industry—An application of survival analysis modeling using SAS. *SAS User Group International (SUGI27) Online Proceedings*, 114.
- Lu, J., & Park, O. (2003). Modeling customer lifetime value using survival analysis—an application in the telecommunications industry. *Data Mining Techniques*, 120-128.

- Martínez, R. M. V. (2004). Gestión de la clientela: la manera de conseguir y retener clientes rentables. *ESIC Editorial*.
- Masarifoglu, M., & Buyuklu, A. H. (2019). Applying survival analysis to telecom churn data. *American Journal of Theoretical and Applied Statistics*, 8(6), 261-275.
- Méndez, R. R., & Santoyo, I. B. (2020). Diferenciación de marca e imagen en la distribución y comercialización de electricidad y gas. *Boletín económico de ICE, Información Comercial Española*, (3121), 51-65.
- MIBEL. (2009). Description of the operation of the MIBEL.
- Neslin, S. A., Grewal, D., Leghorn, R., Shankar, V., Teerling, M. L., Thomas, J. S., & Verhoef, P. C. (2006). Challenges and opportunities in multichannel customer management. *Journal of service research*, 9(2), 95-112.
- Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2), 2592-2602.
- Olivares Gallardo, A. (2014). Libre mercado y regulación: La experiencia en el sector eléctrico español. *Revista chilena de derecho*, 41(1), 205-228.
- OMIE: Qué es, cuáles son sus funciones y más. (n.d.). Plena-energia.com. Recuperado el 02 de mayo de 2023, from <https://www.plena-energia.com/post/omie>
- Pampel, F. C. (2020). Logistic regression: A primer (No. 132). *Sage publications*.
- Peláez, I. M. (2016). Modelos de regresión: lineal simple y regresión logística. *Revista Seden*, 14, 195-214.
- Pérez, C. J. (2016). Guia rapida de Gretl.
- Pérez de Arévalo Trillo-Figueroa, C. (2021). La liberalización del mercado eléctrico español: agentes intervinientes, distribución de competencias y la especial posición del regulador. Retos a los que se enfrenta como consecuencia de la crisis provocada por la Covid-19. *Boletín mexicano de derecho comparado*, 54(161), 733-770.

- Portela, N. F., & Utray, J. F. (2009). El diseño de mercado para el sector eléctrico español. *Papeles de economía española*, (121), 141-158.
- Pribil, J., & Polejova, M. (2017). A churn analysis using data mining techniques: Case of electricity distribution company. In *Proceedings of the World Congress on Engineering and Computer Science* (Vol. 1, pp. 1-6).
- Protocol, K. (1997). Kyoto protocol. *UNFCCC Website*.
- Redondo Morais, J. (2015). Análisis comparativo de alternativas comerciales en el mercado minorista de electricidad para el pequeño consumidor (Master's thesis).
- Rodríguez, C. P., & García, M. R. (2017). La función redistributiva en los servicios de interés económico general. *Economía UNAM*, 14(40), 101-120.
- Rodríguez, D. R. (2021). Sobre los costes, los precios y el mercado de la electricidad. FEDEA, Estudios sobre la Economía Española, (2021-28).
- Rodríguez García, R. (2020). Impacto del Covid-19 en el sector eléctrico.
- Roldán-Fernández, J. M., Burgos-Payán, M., Trigo-García, Á. L., Díaz-García, J. L., & Riquelme-Santos, J. M. (2014, May). Impact of renewable generation in the Spanish Electricity Market. In *11th International Conference on the European Energy Market (EEM14)* (pp. 1-5). IEEE.
- Rosenbaum, E. F. (2000). What is a market? On the methodology of a contested concept. *Review of social economy*, 58(4), 455-482.
- Rozas González, M. P. (2016). Subcontratación en el sector energético.
- Sánchez Rodríguez, R. (2022). ¿Por qué sube el recibo de la luz? Un análisis del mercado eléctrico en España.
- Shaaban, E., Helmy, Y., Khedr, A., & Nasr, M. (2012). A proposed churn prediction model. *International Journal of Engineering Research and Applications*, 2(4), 693-697.
- TECH Universidad Tecnológica España. (2021). Tecnologías y costes de la generación eléctrica. Recuperado 02 de mayo de 2023, de <https://www.techtute.com/ingenieria/blog/tecnologias-costes-generacionelectrica>

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1-9.

Velázquez Gómez, G. (2018). Mercados energéticos en España.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European journal of operational research*, 218(1), 211-229.

Vives, X. (2006). El reto de la competencia en el sector eléctrico. *IESE Occasional Paper*, 6(13), 1-16.

6. Anexos

Anexo I

Código final para la elaboración del modelo de regresión logística múltiple.

```
# preparación de las variables dicotómicas y estandarización de las variables numéricas
datos$Producto <- as.factor(datos$Producto)
datos$Gestor <- factor(datos$Gestor,levels = c( "GestorInterno","Canal"))
datos$Rotacion <- factor(datos$Rotacion)
datos$Churn <- factor(datos$Churn,levels = c("Churn", "NoChurn") )
datos$CENSOR <- factor(datos$CENSOR)
datos$Churn <- relevel(datos$Churn, ref = "NoChurn")
datos$Consumo<-rescale(datos$Consumo)
datos$Duración<-rescale(datos$Duración)
summary(datos)
# gráficas de las variables
# categóricas
par(mfrow = c(1, 4)) #una fila tres columnas
plot(datos$Gestor, main = "Gestor")
plot(datos$Rotacion,main = "Rotación")
plot(datos$Churn,main = "Churn")
plot(datos$Producto,main = "Producto")
# numéricas
par(mfrow = c(1, 3))
hist(datos$Consumo,main = "Consumo estimado (kW)",xlab = "",ylab = "")
hist(datos$Duración,main = "Duración (días)",xlab = "",ylab = "")
hist(datos$Precio,main = "Precio medio OMIE",xlab = "",ylab = "")
# fijación de una semilla inicial antes de un proceso aleatorio
set.seed(1234)
# separación de datos en 80% entrenamiento y 20% prueba
indice_particion1 <- createDataPartition(datos$Churn, p = 0.8, list = FALSE)
train80 <- datos[indice_particion1, ]
test20<- datos[-indice_particion1, ]
# definición del número de divisiones de los datos para la validación cruzada
control <- trainControl(method = "cv", number = 5)
# creación del modelo de regresión logística múltiple
modelo <- train(Churn~Gestor*Rotacion+Duración+Consumo+Producto, data = train80, method =
"glmnet", trControl = control)
# Predicción
```

```
p<-predict(modelo, test20)
matriz_confusion <- confusionMatrix(p, test20$Churn)
print(matriz_confusion)
#resultados
modelo_best_coef <- coef(modelo$finalModel, s = modelo$bestTune$lambda)
modelo_coef <- coef(modelo$finalModel)
print(modelo_coef)
```